

Durham Research Online

Deposited in DRO:

23 October 2012

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Schultz, J. and Pilz, K.S. (2009) 'Natural facial motion enhances cortical responses to faces.', *Experimental brain research.*, 194 (3). pp. 465-475.

Further information on publisher's website:

<http://dx.doi.org/10.1007/s00221-009-1721-9>

Publisher's copyright statement:

The original publication is available at www.springerlink.com

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Natural facial motion enhances cortical responses to faces

Johannes Schultz^{1,*} & Karin S. Pilz^{2,*}

¹ Dept. of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

² Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada

* Both authors contributed equally to this work

Corresponding author: Karin S. Pilz, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, ON, Canada, Tel.: +1-905-525-9140 x24489, Fax: +1-905-529-6225, Email: pilzk@mcmaster.ca

Keywords: facial motion, face localizer, STS, biological motion, FFA, OFA, fMRI.

Abstract

The ability to perceive facial motion is important to successfully interact in social environments. Previously, imaging studies have investigated neural correlates of facial motion primarily using abstract motion stimuli. Here, we studied how the brain processes natural non-rigid facial motion in direct comparison to static stimuli and matched phase-scrambled controls. As predicted from previous studies, dynamic faces elicit higher responses than static faces in lateral temporal areas corresponding to hMT+/V5 and STS. Interestingly, analyses of individually-defined, static-face-sensitive regions in bilateral fusiform gyrus and left inferior occipital gyrus also respond more to dynamic than static faces. These results suggest integration of form and motion information during the processing of dynamic faces even in ventral temporal and inferior lateral occipital areas. In addition, our results show that dynamic stimuli are a robust tool to localize areas related to the processing of static and dynamic face information.

Introduction

Being required to understand and predict the actions of others to be able to successfully interact in a social environment has led our visual system to become particularly sensitive to human movements (for a recent review, see Blake and Shiffrar 2007). Facial motion in particular is a very important cue to judge other people's actions, emotions and intentions towards us (Bassili 1976; Kamachi et al. 2001). In addition to this, facial motion has also been shown to facilitate face recognition (O'Toole et al. 2002; Pilz et al. 2006). Due to the familiarity and behavioral significance of facial motion, it is most likely that our visual system has developed mechanisms that facilitate its perception and it is also very plausible to assume that certain mechanisms exist that integrate invariant and changeable properties of faces (Haxby et al. 2000).

Studies of biological motion, including faces, suggest that the interpretation of the movements and actions of others recruit specialized neural pathways (Allison et al. 2000; Giese and Poggio, 2003; Blakemore et al. 2001). In monkeys, neurons in the anterior part of the superior temporal polysensory area (STPa) were found to respond both to the form and the motion of bodies and heads, indicating integration of form and motion information in this area (Oram and Perrett 1996). In humans, involvement of the superior temporal sulcus (STS) in the processing of relevant and familiar types of biological motion has also been shown, e.g., in response to human body motion (tested using point-light displays, Bonda et al. 1996; Grossman et al. 2000), or to facial motion due to speech production (Campbell et al. 2001; Hall et al. 2005), expression of emotions (LaBar et al. 2003; Pelphrey et al. 2007) or in complex scenes such as movies (Bartels and Zeki 2004; Hasson et al. 2004). Additionally, these regions have been shown to respond to natural images of implied facial motion (Puce et al. 1998; Puce et al. 2003), as well as to natural images of implied body motion (Jellema and Perrett 2003).

Most of the studies investigating the neural correlates of facial motion have used abstract motion stimuli like implied motion from static images (Puce et al. 1998; Puce et al. 2003), moving avatars (i.e. cartoon faces, for example Pelphrey et al. 2005; Thompson et al. 2007), or motion stimuli that were produced by morphing a static towards an emotional face (LaBar et al. 2003; Sato et al. 2004; Pelphrey et al. 2007). Using such ‘unnaturally’ moving stimuli might not fully capture the mechanisms underlying the processing of natural facial motion. The controlled fMRI studies of facial motion that have used video sequences of natural facial motion either focused on differences between types of face motions and thus did not use non-face control stimuli (Campbell et al. 2001; Hall et al. 2005). A recent study by Fox et al., (2008) investigated differences in brain activation between static and dynamic stimuli using non-face stimuli as controls. They applied two localiser scans, one contrasting static images of faces and objects, the other one contrasting dynamic videos of faces and objects. Comparing these two localisers, their results suggest that dynamic localisers are more reliable and more selective than static localisers. Although this study showed the usefulness of using dynamic stimuli to localize areas related to face processing, they were not able to directly compare brain activation towards static and dynamic stimuli, because those stimuli were used in different scanning sessions. Here, we investigated brain activation in response to natural non-rigid face motion and directly compared it to static faces and non-face controls, which is necessary to demonstrate how the face-processing system responds to dynamic as compared with static faces irrespective of low-level cues. We showed observers video sequences of angry and surprised faces, as well as static stimuli of the same emotions. As controls for low-level stimulus properties including motion, we used the phase-scrambled versions of both kinds of stimuli.

Materials and Methods

Participants

Ten observers (4 females, 6 males) from the Tübingen community volunteered as subjects for 12€ per hour. All observers were naïve as to the purpose of the current experiment and had no history of neurological or psychiatric illnesses. All participants provided informed consent and filled out a standard questionnaire approved by the local ethics committee for experiments involving a high field MR scanner to inform them of the necessary safety precautions.

Stimuli

We used video recordings of the face of three male and five female human actors, taken from the Max-Planck database of moving faces (Pilz et al. 2006). For these recordings, each face made two expressive gestures in separate videos: Surprise and anger. The movie clips used in the dynamic face condition (*dynamic faces*) were composed of 26 frames, presented at a frame rate of 25 frames per sec for a total duration of 1040 ms. Figure 1 shows an example of all 26 frames of a video sequence (top panel). The movie clips started with a neutral expression and ended with the peak of the expression in the last frame. The static face images used in the static face condition (*static faces*) were the last frame of each video sequence and thus showed the peak of each expression; each static face was presented for 1040 ms. All stimuli were embedded in a background that consisted of white noise applied to every RGB color channel. For the dynamic stimuli, the same noise was applied to all the frames of the movie, i.e., the background was static.

As control stimuli, we generated phase-scrambled versions of dynamic (*dynamic scrambled*) and static (*static scrambled*) faces. Researchers have often used objects or fragmented face images as a comparison to face images to investigate areas related to face processing (Kanwisher et al. 1997; Kanwisher et al. 1998). We decided to use phase-

scrambled versions of our stimuli as controls, because fragmented images are constituted more of higher spatial frequencies, resulting from the cardinal axes (i.e., edges) that are produced by dividing a relatively smooth picture like a face into randomly rearranged squares (Sadr and Sinha 2004). Phase-scrambled stimuli have been used successfully in recent neuroimaging studies (Eger et al. 2004; Kovacs et al. 2006; Jacques and Rossion 2007; Rousselet et al. 2007). It has been shown that, especially for face recognition, the frequencies around 8-16 cycles across the face are particularly important (Costen et al. 1996; Näsänen 1999; Morrison and Schyns 2001). Spatial frequencies also seem to interact with the recognition of previously learned static and dynamic images (Pilz et al. 2008), suggesting that they contain important information about the identity of the face. In addition, it has been shown that the FFA processes high and low spatial frequencies differently (Vuilleumier et al. 2003; Gauthier et al. 2005; Rotshtein et al. 2007). Using fragmented images as a contrast would have changed our results as a function of spatial frequency content in the phase-scrambled images. Therefore, it was of high importance to preserve the frequency structure of our original stimuli. Furthermore, we wanted to use a type of control stimuli that worked equally well for both dynamic and static faces in controlling for their respective low-level stimulus properties. Phase-scrambling is ideal, because its effect on both static and dynamic faces is very comparable (keeping the spatial frequency content constant while eliminating recognizable shapes).

Phase-scrambling of our images was accomplished as follows. For each independent RGB color channel, the images were transformed into amplitude and phase components using the Fourier transform. Noise patterns were generated by inverse Fourier transform of the original amplitude spectrum of the image but with a random phase spectrum. For the movies, the same random phase spectrum was used for each frame of a given movie but the

amplitudes were those of the original frames. This resulted in control movies that were not flickering.

 Insert Figure 1 about here

Design and Procedure

There were 5 conditions in the experiment: *fixation*, *static faces*, *static scrambled*, *dynamic faces*, and *dynamic scrambled*. The observer's task was a one-back matching task, i.e., they had to press a button whenever two identical stimuli sequentially appeared on the screen. We used a block design with 24 blocks, each composed of 6 stimuli which were presented every 3 seconds. Blocks were history-matched, i.e., every condition was preceded by each condition equally often. Given that there were 16 different face stimuli in total (8 identities times 2 expressions) and 6 stimuli per block, the probability of a stimulus repetition was about 0.31 per block; i.e., each subject would on average encounter about 6 targets distributed across conditions.

Observers lay supine on the scanner bed. The stimuli were back projected onto a projection screen situated behind the observers' head and reflected into their eyes via a mirror mounted on the head coil. The projection screen was 140.5 cm from the mirror, and the stimuli subtended a maximum visual angle of approximately 9.0° (horizontal) x 8.3° (vertical). A JVC LCD projector with custom Schneider-Kreuznach long-range optics, a screen resolution of 1280 pixels x 1024 pixels and a 60 Hz refresh rate were used. The experiment was run on a 3.2 GHz Pentium 4 Windows PC with 2GB RAM and an NVIDIA GeForce 7800 GTX graphics card with 256 MB video RAM. The program to present the

stimuli and collect responses was written in Matlab using the Psychtoolbox extensions (<http://www.psychtoolbox.org>) (Brainard 1997; Pelli 1997). We used a magnet-compatible button box to collect subjects' responses (The Rowland Institute at Harvard, Cambridge, USA).

Image Acquisition

All participants were scanned at the MR Centre of the Max Planck Institute for Biological Cybernetics, Tübingen, Germany. All anatomical T1-weighted images and functional gradient-echo echo-planar T2*-weighted images (EPI) with BOLD contrast were acquired on a Siemens TIM-Trio 3T scanner with an 8-channel phased-array head coil (Siemens, Erlangen, Germany). The imaging sequence for functional images had a repetition time of 1920 ms, an echo time of 40 ms, a flip angle of 90°, a field of view of 256 mm x 256 mm and a matrix size of 64 pixels x 64 pixels. Each functional image consisted of 27 axial slices. Each slice had a thickness of 3.0 mm x 3.0 mm x 2.5 mm with a 0.5 mm gap between slices. Volumes were positioned to cover the whole brain based on the information from a 13-slice parasagittal anatomical localizer scan acquired at the start of each scanning session. For each observer, between 237 and 252 functional images were acquired in a single session lasting approximately 7.5 min, including a 8 sec blank period at the beginning of the run. The first four of these images were discarded to allow for equilibration of T1 signal. A T1-weighted anatomical scans was acquired after the functional runs (MPRAGE; TR = 1900 msec, TE = 2.26 msec, flip angle = 9°, image matrix = 256 mm [Read direction] x 224 mm [Phase], 176 slices, voxel size = 1x1x1 mm, scan time = 5.59 min).

fMRI Data Pre-processing

Prior to any statistical analyses, the functional images were realigned to the first image and resliced to correct for head motion. The aligned images were then normalized into a standard EPI T2* template with a resampled voxel size of $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm} = 27 \text{ mm}^3$ (Friston et al. 1995a). Spatial normalization was used to allow group statistics to be performed across the whole brain at the level of voxels (Ashburner and Friston 1997; Ashburner and Friston 1999). Following normalization, the images were convolved with an 8 mm full width at half maximum Gaussian kernel to spatially smooth the data. Spatial smoothing was used in this study because it enhances the signal-to-noise ratio of the data, permits the application of Gaussian random field theory to provide for corrected statistical inference (Friston et al. 1996) and facilitates comparisons across observers by compensating for residual variability in anatomy after spatial normalization, thus allowing group statistics to be performed.

fMRI Statistical Analyses

Pre-processed fMRI data were analyzed using the general linear model framework implemented in the SPM2 software package from the Wellcome Department of Imaging Neuroscience (www.fil.ion.ucl.ac.uk/spm). A two-step mixed-effects analysis was used, as is common in SPM for group analyses (Friston et al. 1999). The first step used a fixed-effects model to analyze individual data sets. The second step used a random-effects model to analyze the group aggregate of individual results, which come in the form of parameter estimates for each condition and each voxel (parameter maps). As these group statistics are performed at the voxel level, the individual parameter maps need to be in the same anatomical format and were thus computed on the normalized data.

For each observer, a temporal high-pass filter with a cut-off of 128 sec was applied to the pre-processed data to remove low-frequency signal drifts and artefacts, and an autoregressive model (AR 1 + white noise) was applied to estimate serial correlations in the

data and adjust degrees of freedom accordingly. Following that, a linear combination of regressors in a design matrix was fitted to the data to produce beta estimates (Friston et al. 1995b) which represent the contribution of a particular regressor to the data.

Whole-Brain Analysis

The GLM applied to the individual datasets contained separate regressors of interest for the 4 experimental conditions (*dynamic faces*, *dynamic scrambled*, *static face*, *static scrambled*) and the *fixation* condition. Two sets of regressors were created in SPM2 for each of these conditions in the following manner. For each condition, we first modeled the onset and duration of each stimulus as a series of delta functions. The series of delta functions was convolved with a canonical haemodynamic response function (HRF) to create a first set of regressors. The HRF was implemented in SPM2 as a sum of two gamma functions. To create a second set of regressors the delta functions were convolved with the first temporal derivative of the HRF. Therefore, there were a total of 10 regressors in the part of the design matrix used to model experimentally-induced effects. In addition, the design matrix included a constant term and six realignment parameters (yaw, pitch, roll and three translation terms). These parameters were obtained during motion correction and used to correct for movement-related artefacts not eliminated during realignment.

Fitting each subject's data to the GLM produced 3D parameter estimate maps for each of our conditions of interest. We imported these single-subject parameter maps into SPM2's ANOVA model to evaluate group statistics (random effects) for the following contrasts: *static faces* vs. *static scrambled*, *dynamic faces* vs. *dynamic scrambled*, *dynamic faces* vs. *static faces* and the interaction: (*dynamic face* > *dynamic scrambled*) > (*static face* > *static scrambled*). The interaction was the most stringent test of differences between dynamic and

static faces as it controls for movement in the stimuli. SPM2 uses the Greenhouse-Geisser correction for non-sphericity in the data.

We thresholded the statistical maps from the ANOVA at $p < 0.0001$, uncorrected, with a minimum cluster size of 5 voxels. At this threshold, all voxels survived correction for multiple comparisons across all the voxels in the brain at $p < 0.05$ (False Discovery Rate FDR, Genovese et al. 2002) and all clusters survived cluster-wise multiple corrections at $p < 0.05$ (Friston et al. 1994).

Figure 2 (activations rendered on inflated brain) was created using the `spm_surfirend` toolbox (<http://spmfirend.sourceforge.net/>) and displayed using Neurolens software (www.neurolens.org) on the inflated template brain from the Freesurfer toolbox (<http://surfer.nmr.mgh.harvard.edu>).

Regions Of Interest Analysis

In addition to our whole-brain, voxel-wise group analysis, we performed analyses on individually-defined face sensitive regions of interest (ROI). These ROIs were identified using the contrast *static faces > static scrambled*, as follows. We searched in each subject's individual GLM analysis for clusters whose peak response was located less than 10 mm away from the peak response of the clusters found in the group ANOVA. The single-subject GLMs were thresholded at the lower $p < 0.001$ uncorrected threshold during this ROI search (1) because we were looking in regions of a-priori interest which had already survived whole-brain correction in the group ANOVA and (2) to increase the likelihood of finding significant clusters in as many of the individual subjects as possible.

After identifying these individual ROIs, we computed their block-averaged response time-courses to each condition, as follows. Raw BOLD signal data were extracted and filtered by removing low frequencies (cutoff = 128 seconds) and movement artefacts (using

the realignment parameters calculated by SPM2), then averaged over voxels in each ROI. For each run of each participant, the time-series were converted into percent signal change from average activity by dividing the signal measured at each time point by the average signal during the run, subtracting 1, and then multiplying by 100. The block-related responses to each condition were then averaged across all participants from 10s before to 30s after each block onset. The signal from the *fixation* condition was then used as a baseline and subtracted from each of the four other conditions. Therefore, the “0” point on the y axis of Figure 3 corresponds to the mean activity in the *fixation* condition across all runs, and positive and negative values respectively represent relative increases and decreases from the mean signal intensity in the *fixation* condition.

In each ROI, group statistics were assessed as follows. For each block of trials, the magnitude of the response to each condition was calculated by averaging the signal time-course in the period between 7.5 and 19 seconds after block onset. The response to *static faces* and *dynamic faces* was then compared using 2-tailed paired-samples *t*-tests over subjects. To assess the robustness of the magnitude effects to differences in low-level stimulus characteristics, these tests were computed again after subtracting from the response time-course to each *faces* condition the response to the matching *phase-scrambled faces* conditions. This effectively tests the following interaction: (*dynamic face* > *dynamic scrambled*) > (*static face* > *static scrambled*). Note: Our ROIs were defined by comparing *static faces* to *static scrambled*, and thus the response to *dynamic faces* (or to *dynamic scrambled*) did not play any role in the definition of these ROIs (i.e., the voxels of our ROI could respond more, less or similarly to dynamic faces compared to static faces). As the way we defined the ROIs did not influence the outcome of the contrasts testing for responses to *dynamic faces* vs. other conditions, it is perfectly valid to statistically compare responses to *static faces* and *dynamic faces* without a-priori biases introduced through the ROI definition

method. In effect, instead of performing a separate localiser *experiment*, we used some of the conditions of our experiment as a localiser *contrast* to define regions in which we subsequently tested other contrasts (Friston et al. 2006).

Results

Whole-Brain Statistics

Clusters of voxels responding more to *static faces* than to *static scrambled* were found in fusiform gyrus (FFG) bilaterally, in inferior occipital gyrus (IOG) bilaterally and in the right STS. Given their anatomical location (see coordinates in Table 1), the clusters in FFG and IOG most likely correspond respectively to the fusiform face areas (FFA, Kanwisher et al. 1997) and the occipital face areas (OFA, Halgren et al. 1999; Gauthier et al. 2000; Hoffman and Haxby 2000). As we did not define these clusters by contrasting faces against objects as was done in the studies defining FFA and OFA, we prefer to use the terms FFG and IOG. Figure 2A shows these results thresholded at $p < 0.0001$ uncorrected (Note: right STS survived the threshold of $p < 0.05$, whole-brain corrected but not $p < 0.0001$ uncorrected and thus does not appear in Figure 2). Clusters of voxels responding more to *dynamic faces* than to *dynamic scrambled* were found bilaterally in the following structures: fusiform gyri (FFG), inferior occipital gyrus (IOG), in the posterior and middle parts of the superior temporal sulcus (STS) extending into middle (MTS) and inferior temporal sulci, including the anatomical location of area hMT+/V5 (Dumoulin et al. 2000), as well as in middle prefrontal (MFG), medial prefrontal and medial orbitofrontal cortex, inferior frontal gyrus (IFG) and posterior cingulate gyrus (see Figure 2B). A higher response to *dynamic faces* than to *static faces* was found bilaterally in STS (extending into middle temporal gyrus and sulcus, MTS), in the inferior temporal sulcus (hMT+/V5), and in a small cluster in the precentral gyrus (see Figure 2C). No areas were found that responded more to static than dynamic faces. The interaction

(*dynamic face > dynamic scrambled*) > (*static face > static scrambled*) yielded significant effects exclusively in bilateral STS (Figure 2D). Details of the peaks of these activations are reported in Table 1.

Insert Table 1 about here

Insert Figure 2 about here

Individual Face-Sensitive Regions Of Interest

We located the following ROIs in 8 to 10 out of our 10 subjects: left and right FFG, left and right IOG, and right STS. As stated in the previous paragraph, FFG and IOG most likely correspond to FFA and OFA respectively (see coordinates in Table 2). As reported in Table 2 and shown in Figure 3, all ROIs except the right IOG responded more to *dynamic faces* than to *static faces* when both conditions were compared with fixation. In addition, right FFG and right STS also showed increased activation for dynamic compared to static faces when both were contrasted with their matched phase-scrambled controls (i.e., (*dynamic faces > dynamic scrambled*) > (*static faces > static scrambled*)). No ROI showed a higher response to static faces than to dynamic faces.

Note: almost identical time-courses were found in fusiform and occipital ROIs identified using the contrast *dynamic faces > dynamic scrambled*, which is an indication of the great overlap between ROIs identified using both methods.

Insert Table 2 about here

Insert Figure 3 about here

Discussion

In this study, we investigated brain activation in response to dynamic face stimuli using natural video sequences of facial motion and directly compared it to activation in response to static face images. Using ROI analyses, we found that in most of the classic face-sensitive areas (bilateral FFG, left IOG and the right STS), the BOLD response to dynamic faces was higher than to static faces. In right FFG and right STS, these effects survived even when controlling for low-level visual properties of the stimuli using matched phase-scrambled controls. In addition, our analyses confirmed that STS is the brain region most sensitive to dynamic faces when controlling for stimulus motion. No clusters of the whole-brain analysis or any ROI showed greater response to static than dynamic faces. Taken together, these results show higher brain activation for dynamic than static faces not only in areas that have been related to the processing of changeable aspects of faces but also in areas that have been previously attached to the processing of invariant aspects of faces, i.e., the processing of facial form rather than facial motion (Haxby et al., 2000). This is particularly interesting given that face recognition, a process thought to involve mainly areas sensitive to invariant aspects of faces, can be facilitated by facial motion (O'Toole et al. 2002; Pilz et al. 2006). These results suggest an integration of form and motion information in a network of areas including STS, as has been proposed in models of the recognition of biological motion (Giese and Poggio, 2003). In addition, our results provide a strong argument for the use of dynamic stimuli to localize areas related to the processing of human faces, supporting an argument put forward by Fox and colleagues (Fox et al. 2008).

Higher BOLD responses to Dynamic than Static Faces

In almost all face-sensitive ROIs, the BOLD response to *dynamic faces* was higher than to *static faces*. This is consistent with previous results directly comparing dynamic and static faces (Kilts et al. 2003; Sato et al. 2004) and with a recent study showing a stronger differential response in these areas between faces and objects when shown in motion rather than statically (Fox et al. 2008). However, the same contrast performed in the whole-brain analysis did not show significant activation in FFG or IOG (except after lowering the threshold to $p < 0.01$ uncorrected; data not shown). This suggests that the analysis done on individually-defined ROIs is more sensitive, which can be due to several reasons. First, the ROIs were identified individually which compensates for the between-subjects variation in location of functionally-defined regions. Second, the much smaller number of tests being performed in ROI analyses compared to testing all voxels in the brain reduces the multiple comparisons problem and allows more sensitive thresholds to be used in ROI analyses (Saxe et al. 2006).

The higher activation we found for dynamic faces are compatible with the idea that more neurons are tuned to these stimuli because they are more familiar and behaviorally relevant stimuli, as has been suggested by several research groups (Kilts et al. 2003; Bartels and Zeki 2004; Pelphrey et al. 2007; Fox et al. 2008); for example, more neurons have been found that are tuned to frontal views as compared to side views of faces, which could be related to the fact that we have greater experience with frontal view faces (Perrett et al. 1998). We encounter moving faces frequently every day when interacting with other people. Therefore, our visual system is probably more familiar with seeing moving than static faces. As a result, more neurons might be sensitive to dynamic than static faces.

However, the additional number of frames present in the dynamic face stimuli lead to two alternative explanations of our findings: First, an explanation on the level of a single population of neurons sensitive to both static and dynamic faces is that neurons responding to

faces might show response adaptation during the presentation of static faces. Because nothing changes during the presentation of a single static face, the neuronal response would be smaller at the end than at the beginning of each trial, as neural activity and the related BOLD signal are known to decrease when there is no stimulus change (Grill-Spector and Malach 2001). This explanation was also put forward recently by others (Fox et al. 2008). Those same neurons might not adapt during the presentation of the dynamic faces, because the face undergoes subtle changes between successive frames shown during each trial. Less neuronal adaptation during dynamic face presentations might therefore lead to higher metabolic demands and thus to the higher BOLD signal we observe. Given the slow dynamics of the BOLD signal, this difference in neuronal adaptation might also account for the bigger difference in BOLD response we observed at the end of the blocks of trials compared to the beginning.

Second, one could propose an explanation on the level of different neuronal populations that each responds to a particular static frame of the dynamic face stimuli. In this case, all these populations would be active during presentation of our dynamic face stimuli, but only a subset of them would respond to our static face stimuli. This mechanism has also been suggested recently by Fox and colleagues (2008). The difference in the number of static-face-sensitive neuronal populations involved would then explain the difference in BOLD signal we observed, without any involvement of neurons sensitive to face motion per se.

Disproving these alternative hypotheses requires the use of control stimuli with the same number of frames as the dynamic stimuli but not perceived as facial motion. These stimuli are very difficult to create, because simply frame-scrambling our movies yields stimuli perceived as strange, unnatural speeded-up motion, and these perceptual effects probably involve unnatural responses of the face-processing system, leading to further

difficulties in experimental design and interpretation. We are currently addressing this question in further experiments.

Interestingly, the difference in response to dynamic and static faces was not only found in the right STS which is known to respond to biological motion and facial motion (e.g., Haxby et al. 2000; Bartels and Zeki 2004; Hasson et al. 2004), but also in the areas classically known to process invariant aspects of the faces: FFG and IOG. A recent study by Fox and colleagues (2008) also reported a greater difference in response to dynamic faces versus dynamic objects in these areas. But in their study, the responses to dynamic and static faces could not be compared directly. As recognition of facial identity is thought to be mainly accomplished by those latter areas, their higher response to dynamic faces might be linked to the increased recognition performance observed for dynamic faces (O'Toole et al. 2002; Pilz et al. 2006). This will have to be investigated further in purposefully-designed experiments.

Our findings constitute evidence that both motion- and form-related areas participate in the processing of dynamic faces and suggest that temporal and spatial aspect of faces seem to be processed in an integrated fashion in higher level visual brain areas. Those findings are particularly interesting given that the different face identities and expressions in the stimulus set were the same for static and dynamic faces, as were their presentation schedule, and that in some ROIs, these effects even survived when the responses to the phase-scrambled control stimuli was subtracted. Therefore, the effects are not related to face identity or expression differences, and are not simply related to the fact that something was moving in the dynamic face blocks or that each trial was composed of a series of different frames.

Other Regions Responding to Dynamic Faces

Contrasting the parameter estimates for *dynamic faces* to those for *dynamic scrambled*, we found, in addition to activations in the face- and motion-sensitive areas discussed above,

higher activation in inferior and middle frontal gyrus (IFG and MFG) as well as medial prefrontal and orbitofrontal cortex and posterior cingulate gyrus. In their recent paper, Fox and colleagues (2008) found similar results by comparing dynamic faces to dynamic objects. Recent neuroimaging studies have shown that the IFG, prefrontal and inferior parietal areas are important for action observation and imitation (Molnar-Szakacs et al. 2005; Vogt et al. 2007). Iacoboni and colleagues (Iacoboni et al. 2005) found that the ventral premotor cortex responds more to actions observed in an action-related context than in the absence of such a context. They suggest that the human mirror system does not only provide an action recognition mechanism, but also constitutes a neural system for coding the intentions of others. This is supported by studies showing impairment in the recognition of emotional stimuli and attribution of personality traits in patients with lesions in frontal cortex (Damasio et al. 1991; Heberlein et al. 2004). Our stimuli show expressive faces that have a high relevance when interacting in social situations. Therefore, it is reasonable that watching dynamic expressive faces activates areas related to processing of emotional stimuli and observing relevant actions of other people.

The posterior cingulate gyrus has been found to respond more to familiar faces, voices and words (Kim et al. 1999; Leveroni et al. 2000; Shah et al. 2001) and shows an increasing response during acquisition of facial familiarity (Kosaka et al. 2003). Its activation when watching dynamic faces might reflect the fact that dynamic faces are more familiar and / or that dynamic faces automatically trigger processes leading to their familiarization.

Conclusion

This study shows that dynamic faces elicit more activation than both static faces or phase-scrambled controls in form-related face-processing areas (FFG and IOG) and in motion-related face-processing areas (STS). These results are consistent with the hypothesis that our

brain contains mechanisms that are especially tuned to dynamic aspects of faces, and further reveal that regions tuned to invariant aspects of faces respond more to dynamic than static faces. In addition, our results show that dynamic stimuli provide an excellent tool for robustly localizing areas related to the processing of facial form and motion information (also shown by Fox et al. 2008).

Acknowledgments

The work was conducted while both authors were employed at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany. The authors would like to thank Heinrich H. Bülthoff for support. *Conflict of Interest:* None declared.

References

- Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. *Trends Cogn Sci* 4: 267-278
- Ashburner J, Friston K (1997) The role of registration and spatial normalization in detecting activations in functional imaging. *Clinical MRI/Developments in MR* 7: 26-28
- Ashburner J, Friston KJ (1999) Nonlinear spatial normalisation using basis functions. *Human Brain Mapping* 7 254-266
- Bartels A, Zeki S (2004) Functional Brain Mapping During Free Viewing of Natural Scenes. *Hum.Brain Mapp.* 21: 75-85
- Bassili JN (1976) Temporal and spatial contingencies in the perception of social events. *J Pers Soc Psychol* 33: 680-685
- Blake R, Shiffrar M (2007) Perception of Human Motion. *Annu Rev Psychol* 58: 47-73
- Blakemore SJ, Frith CD, Wolpert DM (2001) The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport* 12: 1879-1884
- Bonda E, Petrides M, Ostry D, Evans A (1996) Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci* 16: 3737-3744
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10: 433-436
- Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS (2001) Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res.Cogn Brain Res.* 12 233-243
- Costen NP, Parker DM, Craw I (1996) Effects of high-pass and low-pass spatial filtering on face identification. *Percept Psychophys* 58: 602-612

- Damasio AR, Tranel D, Damasio H (1991) Somatic markers and the guidance of behavior: Theory and preliminary testing. Oxford University Press, New York
- Dumoulin SO, Bittar RG, Kabani NJ, Baker CL, Jr., Le Goualher G, Bruce PG, Evans AC (2000) A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cereb Cortex* 10: 454-463
- Eger E, Henson RNA, Driver J, Dolan RJ (2004) BOLD Repetition Decreases in Object-Responsive Ventral Visual Areas Depend on Spatial Attention. *J Neurophysiol* 92: 1241-1247
- Fox CJ, Iaria G, Barton JJS (2008) Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping Online* in advance of print
- Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS (1995a) Spatial registration and normalisation of images. *Hum Brain Mapp* 2: 165-189
- Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996) Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage* 4: 223-235
- Friston KJ, Holmes AP, Worsley KJ (1999) How Many Subjects Constitute a Study? *Neuroimage* 10: 1-5
- Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995b) Statistical parametric mapping in functional imaging: a general linear approach. *Hum Brain Mapp* 2: 189-210
- Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RNA (2006) A critique of functional localisers. *Neuroimage* 30: 1077-1087
- Friston KJ, Worsley KJ, Frackowiak R, Mazziotta J, Evans AC (1994) Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1: 210-220

- Gauthier I, Curby KM, Skudlarski P, Epstein RA (2005) Individual differences in FICA activity suggest independent processing at different spatial scales. *Cogn Affect Behav Neurosci* 5: 222-234
- Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* 3: 191-197
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15: 870-878
- Giese MA, Poggio T (2003) Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4: 179-192
- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol* 107: 293-321
- Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R (2000) Brain areas involved in perception of biological motion. *J Cogn Neurosci* 12: 711-720
- Halgren E, Dale AM, Sereno MI, Tootell RBH, Marinkovic K, Rosen BR (1999) Location of human face-selective cortex with respect to retinotopic areas. *Hum Brain Mapp* 7: 29-37
- Hall DA, Fussell C, Summerfield AQ (2005) Reading Fluent Speech from Talking Faces: Typical Brain Networks and Individual Differences. *Journal of Cognitive Neuroscience* 17: 939-953
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* 303: 1634
- Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci.* 4: 223-233

- Heberlein AS, Adolphs R, Tranel D, Damasio H (2004) Cortical Regions for Judgments of Emotions and Personality Traits from Point-light Walkers. *J. Cogn. Neurosci.* 16: 1143-1158
- Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci* 3: 80-84
- Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G (2005) Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology* 3: e79
- Jacques C, Rossion B (2007) Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *Neuroimage* 36: 863-876
- Jellema T, Perrett DI (2003) Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia* 41: 1728-1737
- Kamachi M, Bruce V, Mukaida S, Gyoba J, Yoshikawa S, Akamatsu S (2001) Dynamic properties influence the perception of facial expressions. *Perception* 30: 875-887
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302-4311
- Kanwisher N, Tong F, Nakayama K (1998) The effect of face inversion on the human fusiform face area. *Cognition* 68: B1-B11
- Kilts CD, Egan G, Gideon DA, Ely TD, Hoffman JM (2003) Dissociable Neural Pathways Are Involved in the Recognition of Emotion in Static and Dynamic Facial Expressions. *Neuroimage* 18: 156-168

- Kim JJ, Andreasen NC, O'Leary DS, Wiser AK, Ponto LL, Watkins GL, Hitchwa RD (1999) Direct comparison of the neural substrates of recognition memory for words and faces. *Brain* 122: 1069-1083
- Kosaka H, Omori M, Iidaka T, Murata T, Shimoyama T, Okada T, Sadato N, Yonekura Y, Wada Y (2003) Neural substrates participating in acquisition of facial familiarity: an fMRI study. *Neuroimage* 20: 1734-1742
- Kovacs G, Zimmer M, Banko E, Harza I, Antal A, Vidnyanszky Z (2006) Electrophysiological Correlates of Visual Adaptation to Faces and Body Parts in Humans. *Cereb Cortex* 16: 742-753
- LaBar KS, Crupain MJ, Voyvodic JT, McCarthy G (2003) Dynamic perception of facial affect and identity in the human brain. *Cereb Cortex* 13: 1023-1033
- Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM (2000) Neural systems underlying the recognition of familiar and newly learned faces. *J Neurosci* 20: 878-886
- Molnar-Szakacs I, Iacoboni M, Koski L, Mazziotta JC (2005) Functional segregation within pars opercularis of the inferior frontal gyrus: Evidence from fMRI studies of imitation and action observation. *Cereb. Cortex* 15: 986-994
- Morrison DJ, Schyns PG (2001) Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychon Bull Rev* 8: 454-469
- Näsänen R (1999) Spatial frequency bandwidth used in the recognition of facial images. *Vision Res* 39: 3824-3833
- O'Toole AJ, Roark DA, Abdi H (2002) Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn Sci* 6: 261-266

- Oram MW, Perrett DI (1996) Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol* 76: 109-129
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10: 437-442
- Pelphrey KA, Morris J, Michelich C, Allison T, McCarthy G (2005) Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb Cortex* 15: 1866-1876
- Pelphrey KA, Morris JP, McCarthy G, LaBar KS (2007) Perception of dynamic changes in facial affect and identity in autism. *Social Cognitive and Affective Neuroscience* 2: 140-149
- Perrett DI, Oram MW, Ashbridge E (1998) Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* 67: 111-145
- Pilz KS, Buelthoff HH, Vuong QC (2008) Learning influences the encoding of static and dynamic faces and their recognition across different spatial frequencies. *Mem Cog* in press
- Pilz KS, Thornton IM, Bulthoff HH (2006) A search advantage for faces learned in motion. *Exp Brain Res* 171: 436-447
- Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci* 18: 2188-2199
- Puce A, Syngienotis A, Thompson JC, Abbott DF, Wheaton KJ, Castiello U (2003) The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *Neuroimage*. 19: 861-869

- Rotshtein P, Vuilleumier P, Winston J, Driver J, Dolan R (2007) Distinct and convergent visual processing of high and low spatial frequency information in faces. *Cereb Cortex* 17: 2713-2724
- Rousselet GA, Husk JS, Bennett PJ, Sekuler AB (2007) Single-trial EEG dynamics of object and face visual processing. *Neuroimage* 36: 843-862
- Sadr J, Sinha P (2004) Object recognition and Random Image Structure Evolution. *Cognitive Sci* 28: 259-287
- Sato W, Kochiyama T, Yoshikawa S, Naito E (2004) Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research* 20: 81-91
- Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: A defense of functional localizers. *Neuroimage* 30: 1088-1096
- Shah NJ, Marschall JC, Zafiris O, Schwab A, Zilles K, Markowitsch K, Fink GR (2001) The neural correlates of person familiarity: a functional magnetic resonance imaging study with clinical implications. *Brain* 124: 804-815
- Thompson JC, Hardee JE, Panayiotou A, Crewther D, Puce A (2007) Common and distinct brain activation to viewing dynamic sequences of face and hand movements. *Neuroimage* 37: 966-973
- Vogt S, Buccino G, Wohlschlager AM, Canessa N, Shah NJ, Zilles K, Eickhoff SB, Freund H-J, Rizzolatti G, Fink GR (2007) Prefrontal involvement in imitation learning of hand actions: Effects of practice and expertise. *Neuroimage* 37: 1371-1383
- Vuilleumier P, Armony JL, Driver J, Dolan RJ (2003) Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat Neurosci* 6: 624-631

Tables

Table 1. Anatomical and statistical details of the peaks of significant activations revealed by the contrasts performed in the ANOVA group analysis. All activations survive correction for multiple comparisons across the whole brain.

Anatomy	Hemisphere	Coordinates	T	Z
		X, Y, Z		
Static faces > static scrambled				
Fusiform gyrus (FFG)	Left	-42, -48, -24	5.72	4.79
	Right	39, -57, -18	5.62	4.73
Inferior occipital gyrus (IOG)	Left	-39, -72, -12	5.65	4.75
	Right	45, -75, -12	5.79	4.84
Superior temporal sulcus (STS)	Right	51, -48, 21	4.98	4.31
Dynamic faces > dynamic scrambled				
Superior temporal sulcus (STS)	Left	-54, -48, 6	8.09	6.07
	Right	50, -36, 0	7.21	5.64
Fusiform gyrus (FFG)	Left	-45, -51, -21	6.51	5.26
	Right	39, -54, -18	5.67	4.75
Inferior occipital gyrus (IOG)	Left	-39, -72, -12	5.27	4.50
	Right	45, -69, -12	5.71	4.79
Middle prefrontal cortex	Left	-39, 30, 3	5.42	4.60
	Right	51, 33, 0	7.28	5.67
Medial orbitofrontal cortex	Right	3, 42, -15	5.45	4.62
Posterior cingulate cortex	Right	6, -54, 33	5.39	4.58
Inferior frontal gyrus	Left	-48, 18, 24	5.05	4.36
	Right	45, 24, 18	5.18	4.45

Superior medial prefrontal gyrus	Left	-6, 51, 30	4.66	4.09
<i>Dynamic faces > static faces</i>				
Superior temporal sulcus (STS)	Left	-54, -58, 6	8.17	6.10
	Right	63, -27, 0	7.13	5.71
hMT+/V5	Left	-51, -69, 9	8.66	6.33
	Right	45, -66, 3	7.35	5.71
Precentral gyrus	Left	-39, -3, 51	4.58	4.04
	Right	54, 0, 51	4.63	4.07
<i>Interaction: (Dynamic faces > dynamic scrambled) > (Static faces > static scrambled)</i>				
Superior temporal sulcus (STS)	Left	-57, -42, 6	5.91	4.14
	Right	66, -27, 0	4.56	4.02

Note: Coordinates indicate local maxima in MNI space. T and Z column respectively indicate T values and Z scores from whole-brain ANOVA analysis.

Table 2. Location of the individually-defined face-sensitive regions of interest and response differences to dynamic vs. static faces.

<i>Structure</i>	<i>Coordinates (X, Y, Z)</i>		<i>N</i>	<i>Dynamic face > static face</i>	
	<i>Mean</i>	<i>STD</i>		<i>Fix</i>	<i>Scram</i>
Left FFG	-42, -51, -22	0.6, 1.8, 1.1	8	2.45*	1.55
Right FFG	43, -54, -19	0.9, 2.1, 0.8	10	3.50*	2.59*
Right STS	53, -51, 18	1.5, 2.3, 1.6	9	6.55***	4.40**
Left IOG	-40, -76, -10	1.2, 2.1, 1.5	10	2.35*	0.77
Right IOG	45, -76, -11	1.2, 2.6, 1.2	10	1.66	0.79

Notes: Coordinates are in MNI space. *N* indicates number of subjects in which each ROI was identified. “*Dynamic face > static face*” columns show 2-tailed paired *t* values. *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.001$. Tests were performed after subtracting response to *fixation* (“*Fix*” column) or to the corresponding phase-scrambled faces stimuli (“*Scram*”), which is equivalent to the interaction test: (*dynamic face > dynamic scrambled*) > (*static face > static scrambled*). Coordinates are in MNI space.

Figure Captions

Figure 1. Example stimulus images. Top: all 26 frames of an example face movie stimulus (*dynamic face*). Bottom: all 26 frames of an example phase-scrambled face movie stimulus (*dynamic scrambled*). In the *static* conditions, only the last frame of each movie was shown, for the same duration as the dynamic stimuli. Stimuli were shown in color.

Figure 2. Results of the whole-brain ANOVA group statistics projected on the surface of an inflated standard structural scan. Panel **A** shows clusters responding more to *static faces* than *static scrambled*. Panel **B** shows clusters responding more to *dynamic faces* than *dynamic scrambled*. Panel **C** shows clusters responding more to *dynamic faces* than *static faces*. Panel **D** shows clusters with a significant interaction effect: (*dynamic faces* > *dynamic scrambled*) > (*static faces* > *static scrambled*). Insets in D show percent signal change from fixation (mean & SEM over subjects) for *static faces* (SF), *static scrambled* (SS), *dynamic faces* (DF) and *dynamic scrambled* (DS) in left and right STS clusters (left and right insets respectively). Maps are thresholded at $p < 0.0001$ uncorrected, but all activations survive whole-brain correction at $p < 0.05$. Gradient bar shows T values.

Figure 3. Time-courses of responses to *static faces*, *static scrambled*, *dynamic faces* and *dynamic scrambled* in individually-defined face-sensitive ROIs (identified by contrasting *static faces* with *static scrambled*). Average time-courses over subjects and SEM are shown.

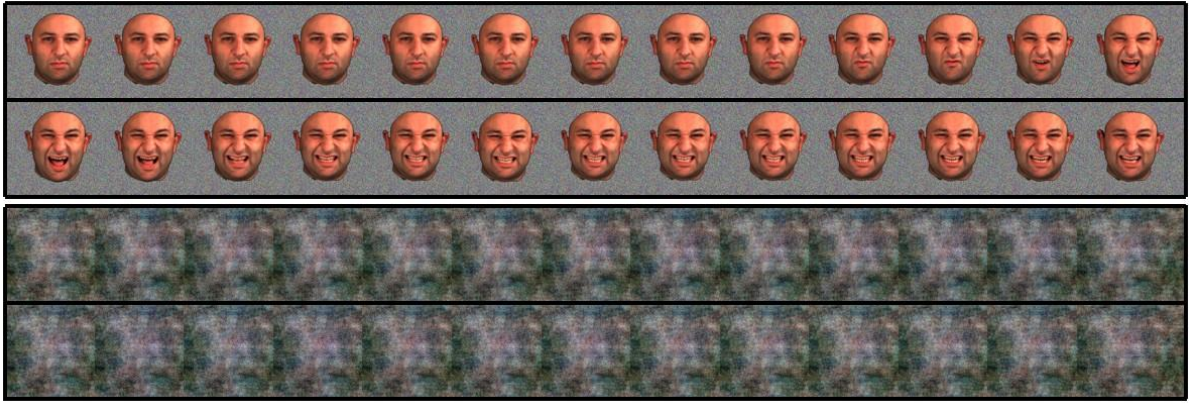


Figure 1

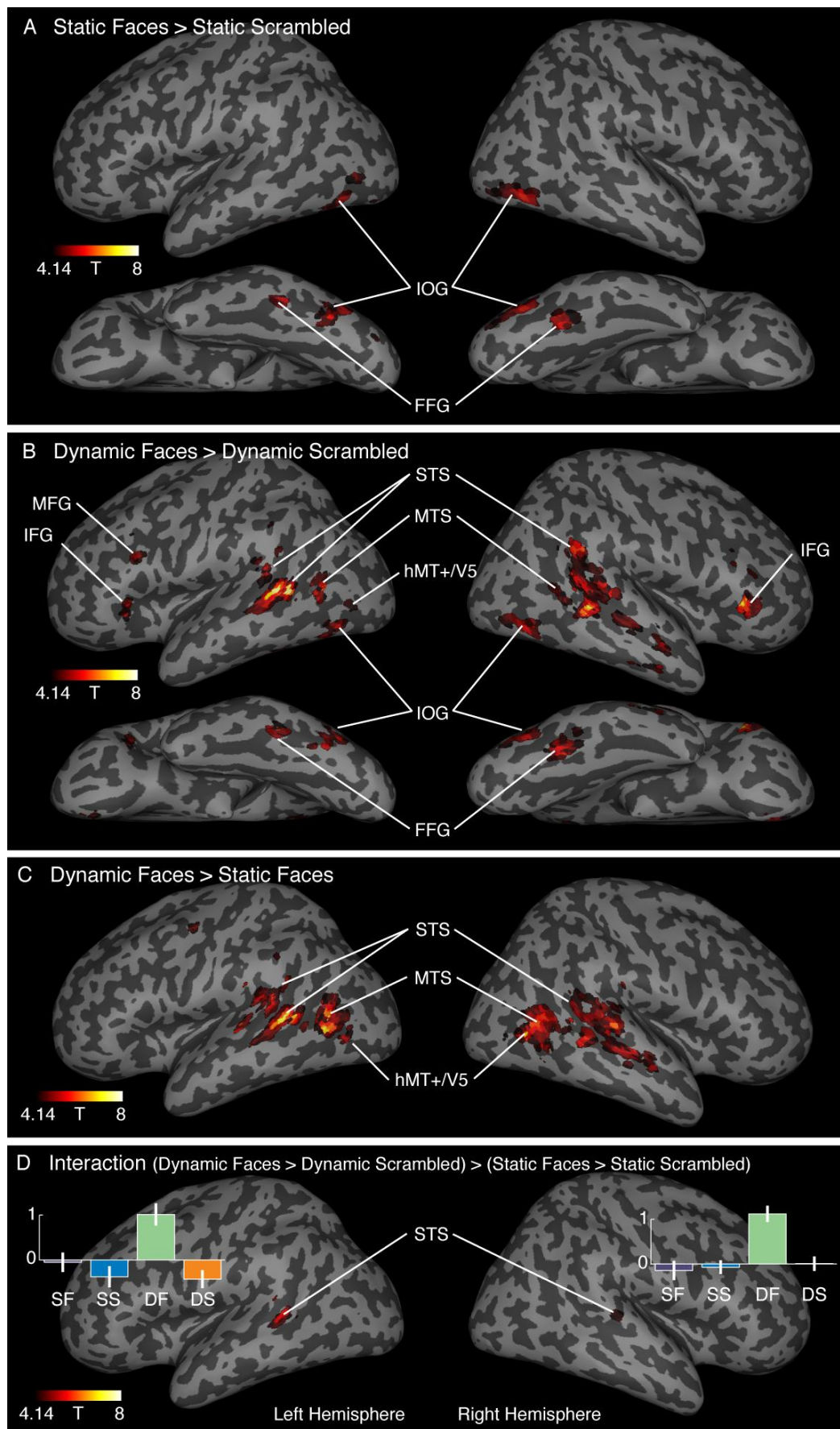


Figure 2

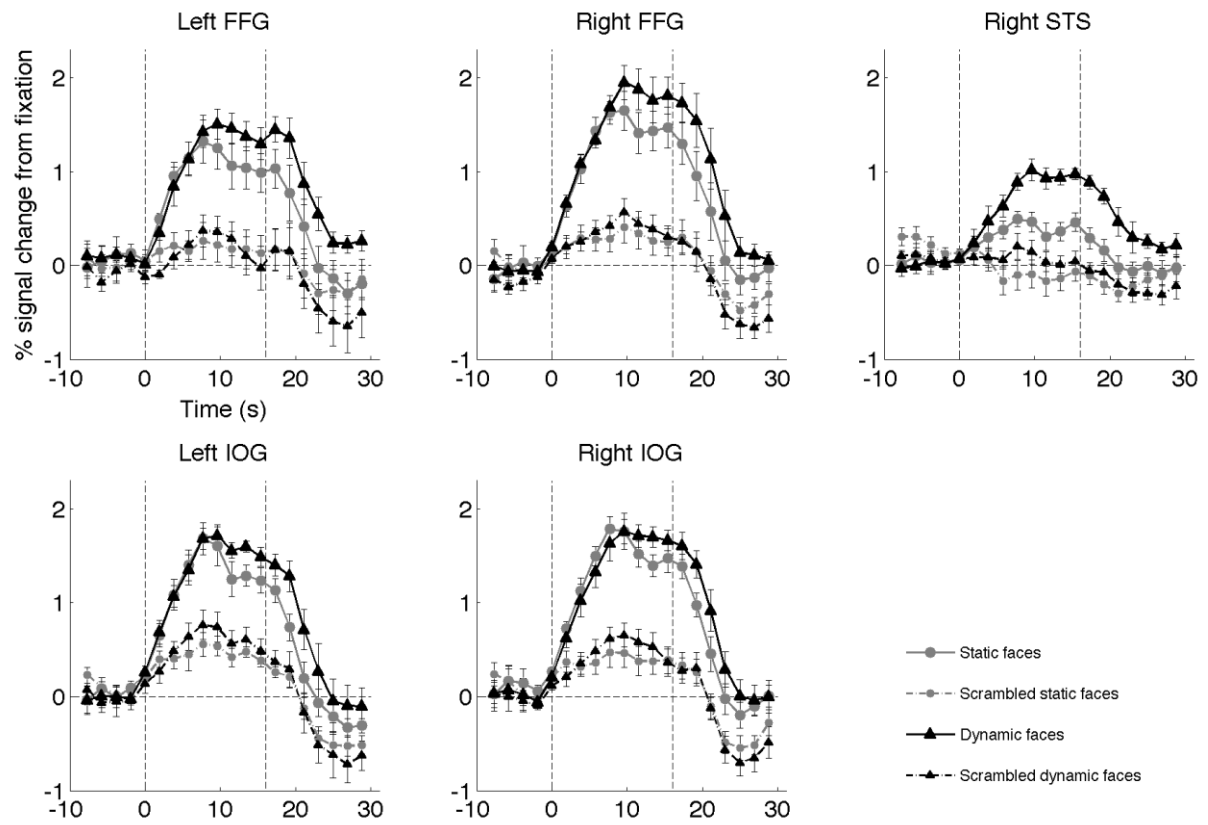


Figure 3